Three-dimensional structure of L-2-haloacid dehalogenase from Xanthobacter autotrophicus GJ10 complexed with the substrate-analogue formate

Ridder, Ivo S.; Rozeboom, Henriëtte J.; Kalk, Kor H.; Janssen, Dick B.; Dijkstra, Bauke W.

Published in:
The Journal of Biological Chemistry

DOI:
10.1074/jbc.272.52.33015

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 52, Issue of December 26, pp. 33015–33022, 1997

Printed in U.S.A.

Three-dimensional Structure of L-2-Haloacid Dehalogenase from *Xanthobacter autotrophicus* GJ10 Complexed with the Substrate-analogue Formate*

(Received for publication, August 22, 1997, and in revised form, October 20, 1997)

Ivo S. Ridder‡, Henriëtte J. Rozeboom‡, Kor H. Kalk‡, Dick B. Janssen§, and Bauke W. Dijkstra‡¶

From the ‡Laboratory of Biophysical Chemistry and the §Laboratory of Biochemistry, Department of Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

The L-2-haloacid dehalogenase from the 1,2-dichloroethane degrading bacterium *Xanthobacter autotrophicus* GJ10 catalyzes the hydrolytic dehalogenation of small L-2-haloalkanoic acids to yield the corresponding D-2-hydroxyalkanoic acids. Its crystal structure was solved by the method of multiple isomorphous replacement with incorporation of anomalous scattering information and solvent flattening, and was refined at 1.95-Å resolution to an *R* factor of 21.3%. The three-dimensional structure is similar to that of the homologous L-2-haloacid dehalogenase from *Pseudomonas* sp. YL (1), but the *X. autotrophicus* enzyme has an extra dimerization domain, an active site cavity that is completely shielded from the solvent, and a different orientation of several catalytically important amino acid residues. Moreover, under the conditions used, a formate ion is bound in the active site. The position of this substrate-analogue provides valuable information on the reaction mechanism and explains the limited substrate specificity of the *Xanthobacter* L-2-haloacid dehalogenase.

The bacterium *Xanthobacter autotrophicus* is capable of growing on short-chain haloalkanes as its sole source of carbon and energy (2). Its natural substrate 1,2-dichloroethane is degraded via 2-chloroethanol, chloroacetaldehyde, and chloroacetate to glycolate in four successive enzymatic reactions before it enters the organism's central metabolic routes. Brominated compounds can also be processed in this way. Two different dehalogenases are used to cleave off the halogen atoms in the first and fourth step. In the first step, a haloalkane dehalogenase catalyzes the conversion of 1,2-dichloroethane into 2-chloroethanol and chloride. The three-dimensional structure (3) and catalytic mechanism (4) of this enzyme have been elucidated by x-ray crystallography.

In the fourth degradation step, a 2-haloacid dehalogenase catalyzes the conversion of chloroacetate to glycolate and chloride. Over 20 different 2-haloacid dehalogenases (EC 3.8.1.2) have been found in various bacteria (5). They have been grouped into four different classes according to their substrate specificity and stereospecific action on 2-monochloropropionic acid: two classes of enzymes that are active with only the L- or D-substrate, yielding products with inversion of configuration at the chiral C-2 carbon atom. The two other classes act on both stereo-isomers, one with inversion of configuration, the other with retention. High amino acid sequence identities are observed among the dehalogenases within any of the separate classes (6, 7), but no homology is evident between the 2-haloacid dehalogenases from different classes.

The 2-haloacid dehalogenase from *X. autotrophicus* belongs to the class of L-specific dehalogenases that act with inversion of configuration (L-DEXs).[1] The *dhlB* gene encoding for it was cloned and sequenced, and the enzyme (DhlB) has been purified and characterized (8). The protein consists of a single polypeptide chain of 253 amino acids and has a molecular mass of 27,431 Da. The amino acid sequence is over 40% identical to seven other L-2-haloacid dehalogenases from various *Pseudomonas* species and one haloacetate dehalogenase from *Moraxella* sp. strain B (9). It shows no homology to haloalkane dehalogenase from *X. autotrophicus* or 4-chlorobenzoyl-coenzyme A dehalogenase from *Pseudomonas* sp. CBS-3 (10), the two other dehalogenases structurally characterized so far.

Dehalogenases act on halogenated aliphatic compounds, which are industrially applied on a large scale as solvents, as intermediates in the production of plastics, and as insecticides. They cause severe problems due to their toxicity as well as their persistence in the environment (11). Since dehalogenases can be used in a biotechnological approach to detoxify these environmentally damaging compounds, they are a fascinating target for research. In addition, the stereospecificity of L-DEXs could make them useful for the biosynthesis of chiral 2-hydroxyalkanoic acids.

^{18}O-Labeled H_2O incorporation experiments have indicated that the dehalogenation reaction catalyzed by L-2-haloacid dehalogenase from *Pseudomonas* sp. YL (L-DEX YL) proceeds via a two-step mechanism, with Asp^{10} acting as a nucleophile in the first step (12). From site-directed mutagenesis, Kurihara *et al.* (9) identified eight other amino acid residues involved in catalysis and substrate binding. Recently, the structure of the L-DEX YL was solved to 2.5-Å resolution (1), which showed that the enzyme has a mixed α/β core domain in a Rossmann-type fold with a four-helix bundle subdomain insertion. The homologous DhlB has 102 out of its 253 amino acid residues in common with the L-DEX YL sequence. To gain more insight into the factors that determine the substrate specificity and stereospecific catalysis, to obtain a structure-based model for

* This work was supported by the Netherlands Foundation for Chemical Research with financial aid from the Netherlands Organization for Scientific Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Laboratory of Biophysical Chemistry, Dept. of Chemistry, University of Groningen, Nÿenborgh 4, 9747 AG Groningen, The Netherlands. Tel.: 31-50363-4381 or 31-50363-4378 (secretary); Fax: 31-50363-4800; E-mail: bauke@chem.rug.nl.

[1] The abbreviations used are: L-DEX, L-2-haloacid dehalogenase; DhlB, L-2-haloacid dehalogenase from *X. autotrophicus* GJ10; L-DEX YL, L-2-haloacid dehalogenase from *Pseudomonas* sp. YL; PEG, polyethylene glycol; bis-Tris, bis[2-hydroxyethyl]imino-tris[hydroxymethyl-]methane; PCMBS, *p*-chloromercuribenzene sulfonate; MIRAS, multiple isomorphous replacement including anomalous scattering; r.m.s., root mean square.

the reaction mechanism, and to complement the data already obtained on the *Pseudomonas* enzyme, we extended our crystallographic studies on dehalogenating enzymes to the L-2-haloacid dehalogenase from *X. autotrophicus*. Here we present the 1.95-Å structure of the enzyme with the substrate-analogue formate bound in the active site, from which a structure-based substrate binding model and the enzyme's reaction mechanism are inferred.

EXPERIMENTAL PROCEDURES

Dynamic Light Scattering—The oligomerization state of L-2-haloacid dehalogenase in solution was determined by dynamic light scattering analysis on a DynaPro 801 instrument (Protein Solutions, Charlottesville, VA), applying 0.1-μm filters for sample injection. A solution of 3.3 mg/ml protein in 100 mM bis-Tris + Tris buffer, pH 7.0, was monitored at 22 °C, with or without the presence of 200 mM sodium formate. Apparent molecular masses of the enzyme as well as the percentage polydispersity were calculated with the instrumentation software using the monomodal assumption. In the absence of formate, the molecular mass fluctuated around 60 kDa (polydispersity of 35%) with deviations up to 20 kDa. The addition of sodium formate reduced the apparent hydrodynamic radius of the molecule, resulting in an estimated molecular mass of 50 kDa (polydispersity 27%) with deviations of about 10 kDa. Subsequent incubation with formate for 1 h lowered the apparent molecular mass to 45 kDa with a corresponding polydispersity of 23%. The fluctuations were reduced to 2–3 kDa and 1–2%, respectively. In a gel filtration experiment with the equilibrated protein solution using a Superdex 75 PC 3.2/30 column (Pharmacia, Sweden), the molecular mass was measured to be 50 kDa.

Crystallization and Heavy Atom Derivative Search—L-2-Haloacid dehalogenase was purified and crystallized as described previously (8, 13). Rod-shaped crystals (typical size $0.4 \times 0.15 \times 0.15$ mm^3) were obtained within a few weeks at room temperature by vapor diffusion in hanging drops using macro-seeding techniques. The drop contained 6 μl 2.5 mg/ml protein, 16% (w/v) PEG 8000, 200 mM sodium formate, and 100 mM bis-Tris buffer, pH 6.8–7.0. The well contained 1 ml of 20–22% PEG 8000, 200 mM sodium formate, and 100 mM bis-Tris buffer, pH 6.8–7.0. The crystals belonged to space group $C222_1$, with cell axes a = 57.2 Å, b = 91.5 Å, c = 84.1 Å for data collected at 120 K (native 1). The crystal volume per unit mass, V_M, was 2.00 Å^3/Da, assuming one molecule in the asymmetric unit. The deduced solvent content was 38%. Under the same conditions also crystals belonging to space group $P2_12_12_1$ were found, with cell axes a = 57.2 Å, b = 84.2 Å, c = 91.5 Å for data collected at 100 K (native 2). Crystals in both crystal forms were of good x-ray quality and diffracted up to 2.0-Å resolution, when synchrotron radiation was used.

Heavy atom derivatives were prepared by soaking the crystals in solutions of a standard mother liquor, containing 25% PEG 8000, 200 mM sodium formate, and 100 mM bis-Tris buffer, pH 6.8–7.0. The pH, heavy atom concentration, and soaking time were varied for promising compounds. In this way, over 25 different compounds were tested under various conditions. X-ray diffraction data were collected at room temperature to check for heavy atom binding. None of the compounds tried yielded a usable derivative, mainly due to problems with non-isomorphism. However, after switching to data collection at cryogenic temperatures, the search readily resulted in four isomorphous derivatives. These were prepared by soaking native crystals for 2 days in 3 mM K_2OsO_4, for 3 days in 1 mM *p*-chloromercuribenzene sulfonate (PCMBS), for 1.5 days in 3 mM ethyl mercurithiosalicylate, and for 4 days in 3 mM $IrCl_3$, respectively. Addition of 25% glycerol to the mother liquor yielded a good cryoprotectant for cryogenic data collection.

Data Collection and Processing—Diffraction data for two native and six derivative data sets were collected in house and at the synchrotron beamlines of the EMBL Outstation at DESY, Hamburg, Germany (see Table I). The in-house data sets (native 1 and the ethyl mercurithiosalicylate and $IrCl_3$) were collected from single crystals at 120 K on an Enraf Nonius FAST area detector system (Enraf Nonius, Delft, The Netherlands). The detector was equipped with a CAD4 κ-goniostat with graphite monochromatized CuKα radiation (λ = 1.5418 Å) from an Elliot GX21 rotating anode x-ray generator. These data were collected and processed using the MADNES package (14). Profile fitting and local scaling of the data was done according to Kabsch (15), and afterward the data were merged with software from the BIOMOL crystallographic package (Protein Crystallography Group, University of Groningen).

All synchrotron data were collected at the beamlines of the EMBL Outstation at DESY, Hamburg, at cryotemperatures (100 K) from single crystals. The osmate and PCMBS derivative data were collected at the X31 beamline, equipped with an 18-cm MAR image plate area detector (MAR Research, Hamburg, Germany), with wavelengths tuned to λ_{Os} = 0.945 Å, and $\lambda_{2\text{-PCMBS}}$ = 1.006 Å, respectively, to make optimal use of anomalous scattering. For the mercury derivative, data sets at $\lambda_{1\text{-PCMBS}}$ = 1.009 Å and $\lambda_{3\text{-PCMBS}}$ = 0.845 Å were also collected to attempt phasing by the method of multiple wavelength anomalous diffraction. A high resolution native data set (native 2) was taken at the BW7B beamline (λ = 0.883 Å), equipped with a 30-cm MAR image plate area detector. However, the crystal used belonged to space group $P2_12_12_1$, which was different from the $C222_1$ crystals observed before. All synchrotron data were integrated and merged with DENZO/SCALEPACK (16). Derivative data were scaled to the native 1 data set with PHASES (17). Data processing statistics are given in Table I.

Structure Determination and Refinement—The structure was initially solved in space group $C222_1$ by the method of multiple isomorphous replacement in combination with anomalous scattering information from all derivatives (MIRAS). All calculations pertaining to the MIRAS analysis and subsequent density modification were done with the PHASES program package. Isomorphous and anomalous difference Patterson maps (10–4 Å data) for the osmate derivative showed a clear single heavy atom site. The resulting phase information was used to calculate isomorphous and anomalous cross-difference Fourier maps for the PCMBS derivative, yielding a single major site. From this, the other two heavy atom data sets could easily be interpreted. For both mercury derivatives, minor sites were identified from difference Fourier maps in the course of the phase refinement. Concerning the three PCMBS data sets, isomorphous differences from only the $\lambda_{3\text{-PCMBS}}$ data were used. Anomalous information from all three PCMBS data sets was included. The final figure of merit was 0.72 for data up to 3.0 Å. Phasing statistics are given in Table I. The MIRAS maps were further improved by solvent flattening, assuming a solvent content of 35%, and by phase extension to 3.0-Å resolution to include calculated intensities for the missing reflections.

This final 3.0-Å resolution electron density map was traced using O (18), and nearly all the chain could be interpreted according to the amino acid sequence (8). One dubious loop connecting two strands was removed after inspection of the L-DEX YL structure that had just been published (1). Molecular replacement routines from AMoRe (19) were then applied to position two copies of the $C222_1$ model in the $P2_12_12_1$ cell. A random set of 5% of the unique reflections was set apart to calculate R_{free} values (20) as an independent validation of the refinement process. The model was subjected to simulated annealing refinement (4000 → 300 K) with X-PLOR (21) with non-crystallographic symmetry constraints. Subsequent cycles of REFMAC (22) and ARP (23) were used in positional and B factor refinement and to automatically place solvent molecules. In the first cycles, tight restraints on non-crystallographic symmetry were applied, but these were gradually released and in later stages the molecules were refined independently since this gave the lowest R_{free} values. At all stages, σ_A-weighted OMIT $2F_o - F_c$ electron density maps (24–26) were calculated and inspected with O to check the agreement of the model with the data. PROCHECK (27) was used to assess the stereochemical quality. Whenever necessary, the model was manually adjusted in O. When the refinement gave no further decrease in R_{free} nor any improvement in stereochemistry, it was considered completed. The initial R_{work} factor was 44.1% (R_{free} = 44.7%) for data in the range of 7–2.0-Å resolution. After refinement, the R_{work} factor had improved to 19.0% for data ranging from 5.0 to 1.95 Å (R_{free} = 24.8%). A final round of refinement including all data resulted in an overall crystallographic R factor of 21.3% for data ranging from 15.0 to 1.95 Å. The final model consists of 3776 protein atoms, 3 formate ions, and 334 water molecules. Refinement statistics are given in Table II. The structure was analyzed using the programs GRASP (28), ASC (29), VOIDOO (30), and programs from the CCP4 suite (31) and the BIOMOL package. The atomic coordinates and structure factor amplitudes have been deposited in the Brookhaven Protein Data Bank with the entry code 1aq6.

RESULTS AND DISCUSSION

Accuracy of the Model—The three-dimensional structure of L-2-haloacid dehalogenase from *X. autotrophicus* was solved by multiple isomorphous replacement supplemented with anomalous scattering information. The refined structure consists of two copies of the molecule, comprising amino acids 1–245 for both monomers. Three formate ions and 334 water molecules are also contained in the structure. For both molecules, the

TABLE I
Data collection and MIRAS analysis

Crystal/derivative	Native 1	Native 2	K_2OsO_4	PCMBS			EMTS[a]	$IrCl_3$
				λ1	λ2	λ3		
Data	X31	BW7B	X31	X31	X31	X31	Fast	Fast
No. of observations	49,346	438,845	20,642	30,247	31,081	52,945	41,312	20,120
Unique reflections	10,198	32,507	5608	4578	4653	7739	4087	4029
Resolution (Å)	2.3	1.95	2.8	3.0	3.0	2.5	3.0	3.0
Completeness (%)	97.0	99.4	98	96	98	98	89	87
R_{sym} on I (%)[b]	6.0	4.2	3.6	2.7	2.8	3.3	7.5	5.8
R_{der} (%)[c]			15.5			13.4	15.0	11.8
MIRAS analysis								
Resolution used (Iso/ano)			3.8/3.0	−/3.0	−/3.0	3.8/3.0	3.0/3.5	3.3/5.0
No. of sites			1	3	3	3	2	1
Phasing power (Iso/ano)[d]			1.0/1.4	−/1.5	−/1.7	1.2/1.6	1.8/1.2	1.1/1.1
Overall figure of merit				0.716				

[a] Ethyl mercurithiosalicylate.
[b] $R_{sym} = \Sigma_h\Sigma_i|I_{(h,i)} - \langle I_{(h)}\rangle|/\Sigma_h\Sigma_i\langle I_{(h)}\rangle \times 100\%$, where $I_{(h,i)}$ is the scaled intensity of the ith observation of reflection h and $\langle I_{(h)}\rangle$ is the mean value.
[c] $R_{der} = \Sigma_h| |F_{PH}| - |F_P| |/\Sigma_h|F_P| \times 100\%$, where F_{PH} and F_P are the structure factors for derivative and native data.
[d] Phasing power = $\langle F_H\rangle/\langle E\rangle$, where $\langle F_H\rangle$ is the root mean square calculated heavy atom structure factor amplitude and $\langle E\rangle$ is the root mean square lack of closure error; isomorphous and anomalous differences are treated separately.

TABLE II
Refinement statistics and stereochemical quality of the final model

	Molecule		Overall
	A	B	
r.m.s. deviations from ideality for			
Bond lengths (Å)	0.009	0.009	0.009
Bond angles (degrees)	2.1	2.0	2.1
Dihedral angles (degrees)	23.8	23.8	23.8
ΔB for bonded atoms (Å^2)	4.9	4.8	4.9
Average B value (Å^2)	21.5	21.0	21.3
Final R factor (5.0–1.95 Å, work set)			19.0%
Free R factor (5.0–1.95 Å, 5% test set)			24.8%
Overall R factor (15.0–1.95 Å, all data)			21.3%
Estimated coordinate error (Å)[a]			0.2

[a] Derived from a Luzzati plot (32).

C-terminal residues 246–253 are not included, because the electron density was too weak to identify their positions unambiguously. In both molecules, the amino acid residue at position 84, a glycine according to the sequence, has clear extra forked electron density extending from Cα, allowing for a Leu, Asn, or Asp. Since it is on the outside of the protein and the $F_o - F_c$ electron density is flat, it can be either an asparagine or an aspartate residue. On the basis of on the DNA sequence (GGG for Gly (8), GA(U/C) for Asp, and AA(A/U) for Asn), we have chosen to model in an aspartate. All other regions of both molecules fit the electron density very well (Fig. 1), as is also reflected in the average real-space correlation coefficient of 0.86 for the final σ_A-weighted OMIT $2F_o - F_c$ electron density map. In two short loops and the C terminus in both molecules (residues 25–27, 204–206, and 244–245), the main chain seems to be disordered and only one of the possible conformations for each loop is represented in the model.

A comparison of the two independently refined dehalogenase molecules in the asymmetric unit shows r.m.s. differences of 0.30 Å on Cα atoms, which is above the coordinate error of 0.2 Å as estimated from a Luzzati plot (32) (data not shown). However, an important part of these differences occurs in the C terminus and the two short loops that seem to be moderately disordered. Another major contribution stems from two short loops, residues 105–107 and 126–132, on the outside of the protein. This is due to their different environment in the two monomers, caused by different crystal contacts in these two regions. Leaving these loops out of the comparison results in an r.m.s. difference of 0.19 Å for 225 superimposed Cα atoms. Thus, the small variations between the two molecules in the asymmetric unit must be attributed to conformational flexibility and/or a different crystal environment of regions on the outside of the dimer rather than to biologically significant structural differences between the monomers. All results discussed below apply to both molecules, unless stated otherwise.

The crystallographic R factor is 21.3% for data in the resolution range of 15–1.95 Å. In a Ramachandran plot (not shown) 93.9% of the residues are in the most favored regions, while the remaining 6.1% lie in the additionally allowed parts. The overall G factor from PROCHECK is +0.10, and the r.m.s. deviations from ideal geometry are 0.009 Å and 2.1° for bond lengths and angles, respectively (33). The average B value for all protein atoms is 20.7 Å^2 and 31 Å^2 for non-protein atoms. Thus, it can be safely concluded that the geometry of the model is sound.

Protein Structure—The L-2-haloacid dehalogenase molecule, with approximate dimensions of 53 × 43 × 38 Å^3, consists of three domains and has an overall shape resembling a heart (Fig. 2). The main domain, comprising amino acids 1–14, 94–192, 220–245, and possibly the C terminus, is composed of a central six-stranded parallel open twisted β-sheet (β1–β6), flanked on both sides by in total five α-helices (α5–α8 and α11). The order of the strands in the β-sheet is β6-β5-β4-β1-β2-β3, with the strands connected by right-handed β-α-β motifs, except for the connection between strands β5 and β6. In the main domain, three 3_{10}-helices are found immediately before and after strand β3 and before strand β4. The folding motif of the main domain is generally known as the Rossmann fold (34).

Two subdomains stand out on the C-terminal end of the central β-sheet: the large atrium of the heart, composed of residues 15–93 and inserted between strand β1 and helix α5, has a distorted anti-parallel four-helix-bundle structure. Helices α1 and α3 are oriented in approximately the same direction, while the 20-residue long helix α2 and helix α4 are at angles of 135° and 225°, respectively. The helices are connected by short loops. Two 3_{10}-helices mark the start and end of this first subdomain. In between the main domain and the large atrium, the active site is located in a cavity with a volume of approximately 80 Å^3. Amino acid residues from the C-terminal end of strands β4, β1, and β2 make up the bottom of the cavity, and residues from the four-helix-bundle domain shield the top of the cavity from the solvent. The second subdomain or small atrium of the heart is made up by residues 193–219, forming an excursion of two anti-parallel helices between strands β5 and β6. This second subdomain is absent in all other L-2-haloacid dehalogenase sequences found to date. The two helical subdomains are not packed tightly to each other, making only one direct hydrogen bond. However, both atria are intimately involved in dimerization (see below).

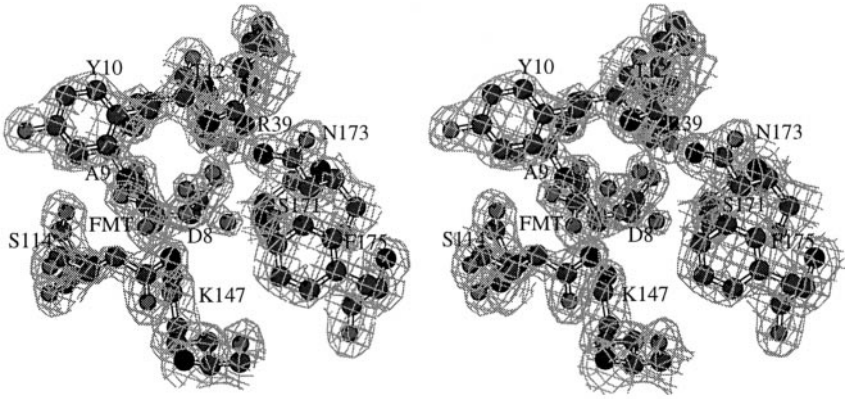

FIG. 1. **Stereo view of the active site with formate ion, displayed with final $2F_o - F_c$ OMIT electron density, contoured at 1σ level.** Figure was generated using BOBSCRIPT (40).

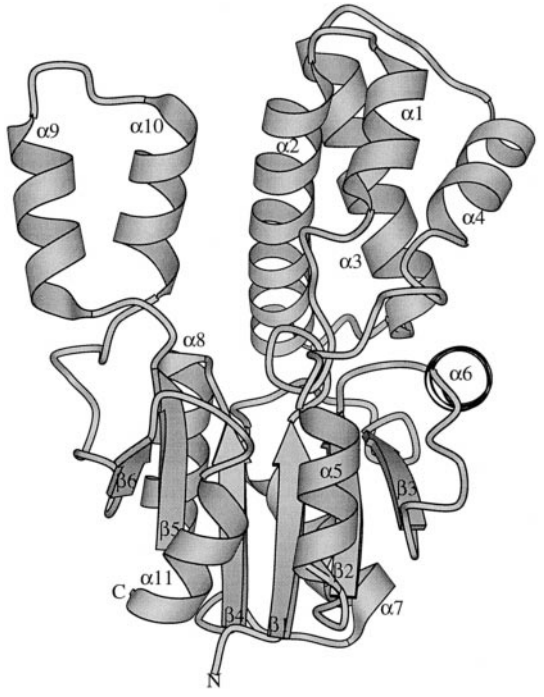

FIG. 2. **Schematic drawing of the structure of L-2-haloacid dehalogenase from *X. autotrophicus*.** Strands are shown as *arrows* and helices as *spirals*, with secondary structure elements and termini labeled. The large atrium comprises helices α1–α4; the small atrium consists of helices α9 and α10 (this and all following figures were generated using MOLSCRIPT (41)).

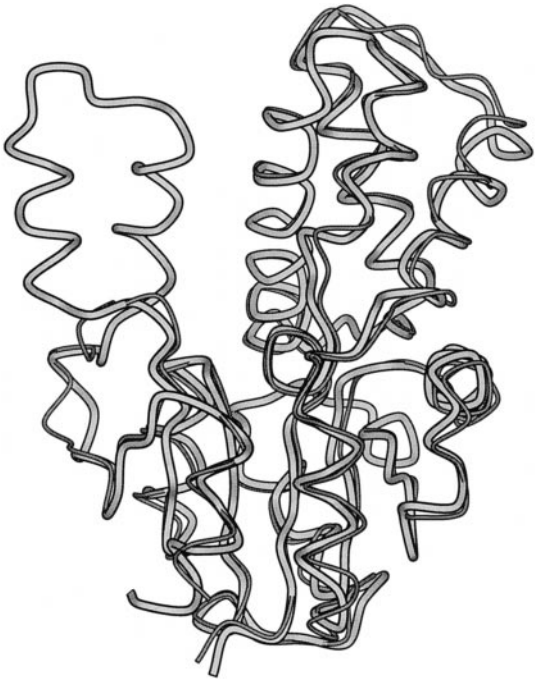

FIG. 3. **Superimposition of L-2-haloacid dehalogenase monomers from *X. autotrophicus* (*thick, light shaded lines*) and *Pseudomonas* sp. YL (*thin, darker shaded lines*).**

Comparison with Other Dehalogenases—Apart from this L-2-haloacid dehalogenase, the family of L-2-haloacid dehalogenases comprises eight other members, one haloacetate dehalogenase from *Moraxella* sp. strain B and seven other L-2-haloacid dehalogenases, all from *Pseudomonas* species, sharing 36–70% amino acid sequence identity. In 1996, the crystal structure of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL was solved (1) to 2.5 Å. The L-DEX YL has 102 out of its 232 amino acid residues in common with the DhlB sequence, and its fold is highly similar; 218 of the 220 Cα atoms can be superimposed with an r.m.s. difference of only 1.3 Å (see Fig. 3). All residues that were identified to be of importance for the reaction mechanism are conserved (9) and can be found in similar positions in the structure of DhlB, although the orientation of the side chains sometimes deviates (see also below). Despite the similarities, two major differences exist, the first being the presence of the left atrium subdomain with the two antiparallel helices α9 and α10. Residues from both helices are part of the dimer interface, which results in a larger contact area than in the L-DEX YL enzyme without changing the overall architecture of the dimer. Second, the active site cavity of the *Xanthobacter* structure is shielded from the solvent, whereas in the *Pseudomonas* sp. YL enzyme it is open from two sides.

One entrance to the active site is via a wide open cleft between the core- and subdomain, which was proposed to accommodate the long alkyl tails of substrates like L-2-bromohexadecanoic acid (1, 35). This entrance is closed in our dehalogenase, due to 1) a different orientation of the Arg^{39} side chain; 2) the position of helix α10, of which residues Phe^{213} to Met^{218} occupy part of space of the L-DEX YL cleft; and 3) a concerted displacement of about 2 Å of the N- and C-terminal ends of the large atrium subdomain, causing the side chains of Tyr^{10}, Asp^{15}, Val^{16}, and Gln^{17} of helix α1 and Tyr^{95} of helix α4 to close in on the cleft. This turns the wide canyon of the *Pseudomonas* YL enzyme into a narrow, blocked-off tunnel (see Fig. 4). The deepest point of the tunnel is 5–6 Å away from the active site cavity.

The second, smaller entrance found in the L-DEX YL structure is located in between helix α4 and the N termini of helices α3 and α6. A narrow tunnel leads from there into the active site cavity, passing along the C termini of strands β1 and β2. In the *Xanthobacter* dehalogenase, this entrance is shut by the side

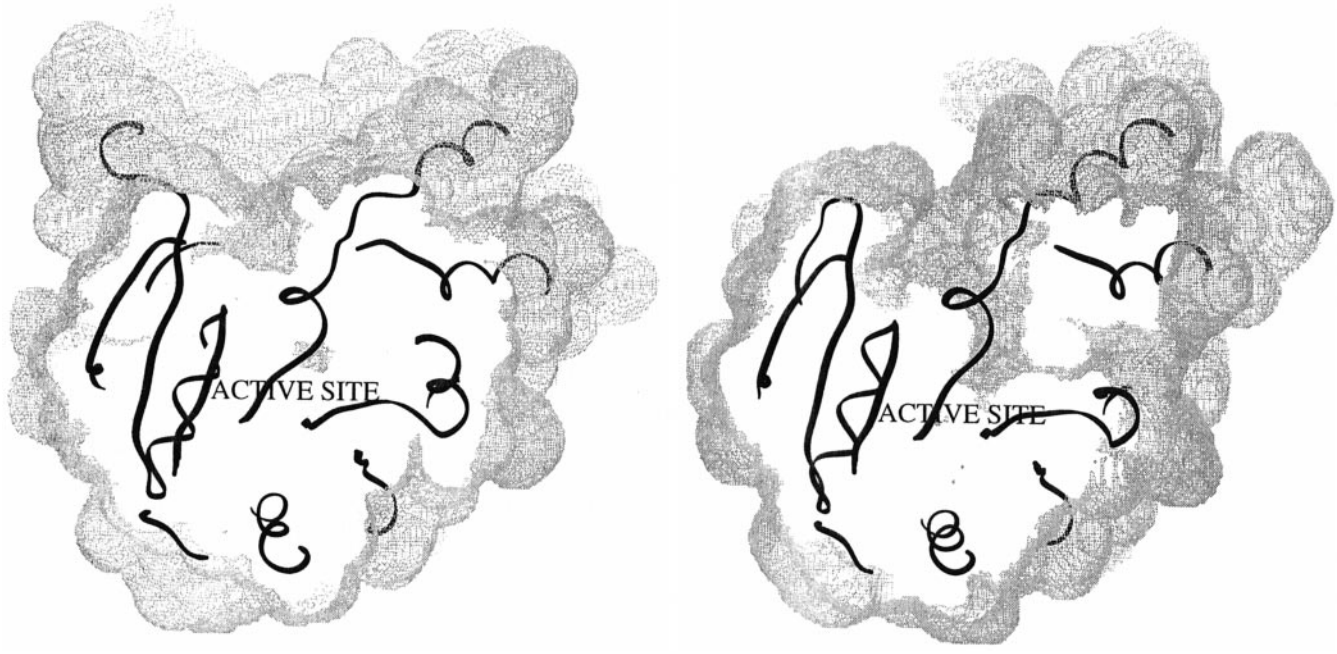

FIG. 4. **Section of the molecular surfaces of DhlB monomer (left) and L-DEX YL monomer (right).** Main chain is represented in *black*, and active site is labeled. The orientation is similar to Fig. 2. The closed active site of DhlB is shielded from the solvent by the small atrium.

chains of Trp^{59} and Met^{120}. The latter makes a sulfur-aromatic interaction with its Sδ to the edge of the tryptophan ring. The corresponding residues in L-DEX YL are a glutamine, which is directed away from the tunnel, and a serine. Moreover, in DhlB residues from helices α3 and α4 fill up the tunnel. The different cavity shapes provide an explanation for the difference in substrate specificity, since the *Xanthobacter* dehalogenase can only efficiently degrade short substrates up to the size of L-2-propionate (8), whereas the *Pseudomonas* YL enzyme is still active on, *e.g.*, L-2-bromohexadecanoic acid. The narrow specificity of DhlB coincides with that of the *Xanthobacter* haloalkane dehalogenase further upstream in the 1,2-dichloroethane degradation route, which is also limited to short substrates.

Structures of two other dehalogenases have been solved to date, the 4-chlorobenzoyl-coenzyme A dehalogenase from *Pseudomonas* sp. CBS-3 (10) and the haloalkane dehalogenase from *X. autotrophicus* (3, 36). The tertiary structure of 4-chlorobenzoyl-CoA dehalogenase consists of two domains, with the active site containing N-terminal main domain comprising two almost perpendicular mixed β-sheets of in total ten strands, surrounded by eight α-helices. A C-terminal domain of three α-helices is involved in trimerization. Thus, its molecular architecture is completely different from L-2-haloacid dehalogenase. The haloalkane dehalogenase bears slightly more resemblance to the L-2-haloacid dehalogenase. It has a main domain made up by a central eight stranded mostly parallel β-sheet surrounded by α-helices with a five helix cap domain on top of it, creating an active site cavity where dehalogenation takes place. At closer inspection, the enzymes turn out to be clearly different; haloalkane dehalogenase is monomeric, the order of the strands in the β-sheet is different, the location of the active nucleophile Asp^{124} is on another strand, and the other two members of the catalytic triad in haloalkane dehalogenase, His^{289} and Asp^{260}, have no counterparts in L-2-haloacid dehalogenase. Although the fold of the three structurally characterized dehalogenases is markedly different, they all dehalogenate their substrates in a two-step mechanism, with an aspartate residue as the active nucleophile in the first step of the reaction (4, 10, 37). This could indicate a general way by which nature copes with the cleavage of carbon-halogen bonds.

Dimerization—The members of the L-DEX family have been reported to occur as homodimers (four enzymes, including L-DEX YL) or homotetramers (one enzyme) (5). To assess the as yet unknown oligomerization state of DhlB, we performed dynamic light scattering experiments, showing that it exists as a dimer in solution at pH 7.0. Addition of formate decreased the polydispersity and stabilized the dimer, explaining the indispensability of the additive in crystallization. Both experimentally determined molecular masses, 45 kDa (dynamic light scattering) and 50 kDa (gel filtration), are close to the calculated molecular mass of 55 kDa for the dimer.

Mimicking the crystallization conditions stabilized the dimer already present in solution. Therefore, we assume that the DhlB crystal structure, with dimensions of $71 \times 43 \times 56$ Å^3 for the dimer, represents the structure of the dimer in solution. Its two-fold symmetry axis runs nearly parallel to the long helices α2 (Fig. 5). The enzyme crystallizes in two different space groups. In the $C222_1$ crystal form, the dimer axis coincides with a crystallographic two-fold axis, in the $P2_12_12_1$ lattice this exact symmetry is lost. However, the crystal packing is highly similar and the monomers have largely maintained their symmetry. A considerable part of the solvent-accessible surface of the protein is buried upon dimer formation. The surface area exposed to solvent amounts to 11,400 Å^2/monomer, whereas the dimer has a solvent-accessible surface of 18,400 Å^2, so 19% of the monomer surfaces is buried. The buried surface is mainly hydrophobic (67%).

In total, 31 residues per monomer make up the dimerization interface, contributed nearly exclusively by the two subdomains of the protein. Residues from helices α2 and α3 in the large atrium and helices α9 and α10 in the small atrium provide most of the contacts. In addition, a few residues from the main domain loop following strand β3 and the N-terminal part of helix α8 are part of the interface. Most of the amino acid residues involved in dimerization are conserved in the family of L-2-haloacid dehalogenases (9) or have conservative mutations, *e.g.* Tyr/Phe^{68}. DhlB dimerizes in a way very similar to L-DEX YL, but the small atrium subdomain, which is absent in all other L-2-haloacid dehalogenases, provides a major contribution to the dimer interface as 75% of the contacts involve at

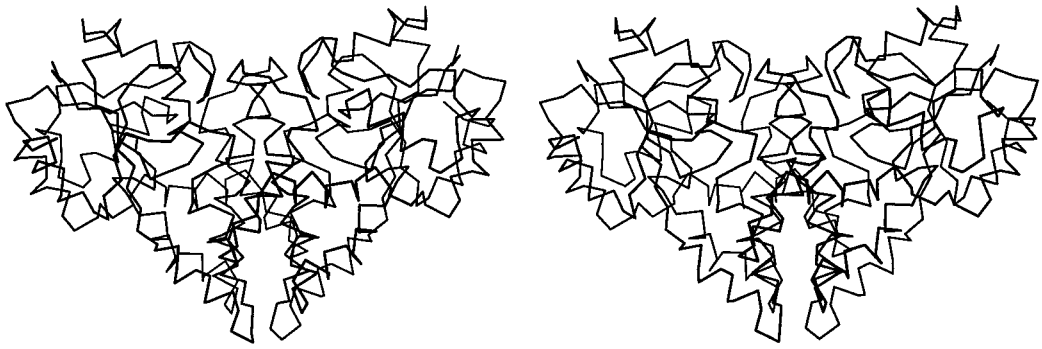

FIG. 5. **Stereo view of the L-2-haloacid dehalogenase dimer.** The two-fold axis relating the monomers is in the vertical direction.

least one residue partner located in this subdomain. Consequently, the *Xanthobacter* dimer is more tightly packed than the L-DEX YL enzyme, where only 13.5% of the solvent-accessible surface is buried upon dimerization.

From the DhlB structure, it is evident that both active sites are accessible from the solvent via a narrow cleft. A functional role for dimerization has not yet been established. Nevertheless, it is remarkable to note that the clefts of both monomers are approximately in line with the two-fold axis relating the monomers. Since at optimum pH the substrate bears a negative charge on its carboxylate group, electrostatic interactions might play a role in pulling the substrate into the active site. Preliminary electrostatic potential calculations using only charges on the side chains show that an electrostatic dipole exists in the dimer. The dipole is directed along the two-fold axis and thereby into the clefts, with the positive pole inside the enzyme. This creates an electrostatic force that could pull substrates in a "carboxylate-first" manner into the active sites of the dimer. In the separate monomers, a dipole is present too, but this has a completely different direction and only upon dimerization is the resulting dipole directed along the two-fold axis for reasons of symmetry. In L-DEX YL an equivalent electrostatic dipole is observed in the dimer structure. These data suggest a role for dimerization in the formation of an electrostatic dipole that favors substrate import.

Active Site—By ^{18}O-labeled H_2O incorporation experiments, Liu *et al.* (12) identified an aspartate residue at the C terminus of strand β1 as the active nucleophile in L-DEX YL. This aspartate is fully conserved in the family of L-2-haloacid dehalogenases, and the equivalent residue in DhlB is Asp^8. In the x-ray structure of *Xanthobacter* L-2-haloacid dehalogenase, it is located at the bottom of the closed active site cavity at the C-terminal end of the β-sheet, in between the main domain and the four-helix-bundle subdomain. The cavity is lined by atoms from Asp^8, Ala^9, Tyr^{10}, Arg^{39}, Leu^{43}, Phe^{58}, Leu^{113}, Ser^{114}, Asn^{115}, Gly^{116}, Lys^{147}, Asn^{173}, Phe^{175}, and Asp^{176} (see Fig. 6. All these amino acids are conserved or show conservative mutations. The geometry of the active site is fixed by an intricate hydrogen-bonding network, involving all the residues around the cavity, except for Phe^{175}, which makes aromatic-aromatic interactions with Phe^{146} and Tyr^{10}. Inside the cavity, one formate ion (see below) and three water molecules are buried. The water molecules make hydrogen bonds to Asn^{173} Nδ2, the guanidinium group of Arg^{39}, the carbonyl oxygen of Tyr^{10}, Asp^8 Oδ1, and with each other.

In haloalkane dehalogenase, the active site is also located in a cavity buried inside the enzyme between the main domain and the cap domain. Nearly all the amino acids in the haloalkane dehalogenase cavity are hydrophobic, with as many as six aromatic residues that surround the cavity, all serving to accommodate the non-polar substrate 1,2-dichloroethane. Only the catalytic residues Asp^{124} and His^{289} are charged. The situation is somewhat different in DhlB, as, apart from Asp^8, two more charged residues (Arg^{39} and Lys^{147}) and several more polar amino acids (Ser^{114}, Asn^{115}, and Asn^{173}) are found with their side chains inside the cavity. The functional role of these residues is discussed below.

Substrate Binding Model and Reaction Mechanism—L-2-Haloacid dehalogenase is a hydrolytic enzyme with Asp^8 as the active site nucleophile (12). Upon attack of the substrate this residue forms an ester intermediate (38), which is then hydrolyzed by a water molecule. A similar mechanism is observed in haloalkane dehalogenase (4), where Asp^{124} is the active nucleophile. The ester intermediate is hydrolyzed by a water molecule that is activated by His^{289}. Together with Asp^{260}, these two amino acids form a classic catalytic triad, also found in other classes of hydrolytic enzymes. From the crystal structure of L-2-haloacid dehalogenase, it can be seen that such a catalytic triad is absent, since no histidine is found in the active site. However, in addition to Asp^8, eight polar and charged amino acids were found to be critical for catalytic activity in the *Pseudomonas* sp. YL enzyme (9). As these residues are fully conserved in the family of L-DEXs, we assume that they have similar functions in all the members of this family.

In the active site of the DhlB structure, a formate ion is bound, which we propose to mimic the carboxylate moiety of the L-2-haloalkanoate substrates (see Fig. 6). The formate C atom is within 3 Å from the Oδ1 of the active nucleophile Asp^8. It is held in position by four hydrogen bonds; the formate O1 is hydrogen-bonded to the backbone amide of Asn^{115} (2.8,Å), and formate O2 makes a weak, forked interaction to the backbone amides of Ala^9 (3.0 Å) and Tyr^{10} (3.1 Å) and a short hydrogen bond interaction with Ser^{114} Oγ (2.6 Å). This explains the importance of the serine for substrate binding affinity, as observed by Kurihara (9). In contrast, the two positively charged residues in the active site cavity, Arg^{39} and Lys^{147} are not involved in binding the substrate's carboxylate group. If the formate is hypothetically extended by one methyl group to an acetate, this methyl group is nicely positioned within 2.5 Å of the Oδ1 of the nucleophile Asp^8, a good distance for nucleophilic attack. The other carboxylate oxygen of Asp^8 is firmly held in position by hydrogen bonds to Thr^{12} Oγ1, Ser^{171} Oγ, and Asn^{173} Nδ2, all three catalytically critical amino acids. These residues might make up the oxyanion hole, stabilizing the negative charge that develops on the Asp^8 Oδ2 upon formation of the tetrahedral intermediate in the hydrolysis step. Ser^{171} Oγ is also within hydrogen-bonding distance of the Asp^{176} side chain, which in turn is oriented by Tyr^{153} Oη. The other carboxylate oxygen of Asp^{176} makes a hydrogen bond to Lys^{147} Nζ. This, together with a hydrogen bond to Leu^{113} O, positions the lysine side chain for a third hydrogen-bonding interaction to Asp^8 Oδ1. The *syn* lone pair electrons of this oxygen are now oriented such that they can perform the nucleophilic attack on the substrate. Finally, the only catalyti-

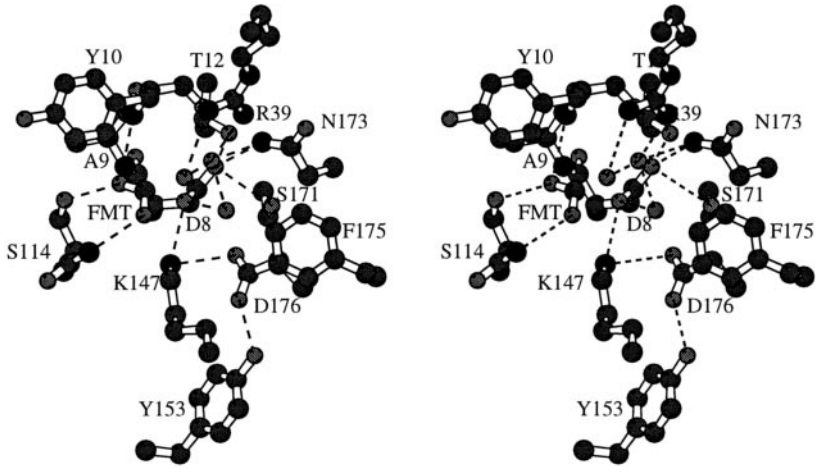

FIG. 6. **Stereo view of the active site with solvent structure and formate ion.** The hydrogen bonds are drawn as *dashed lines*.

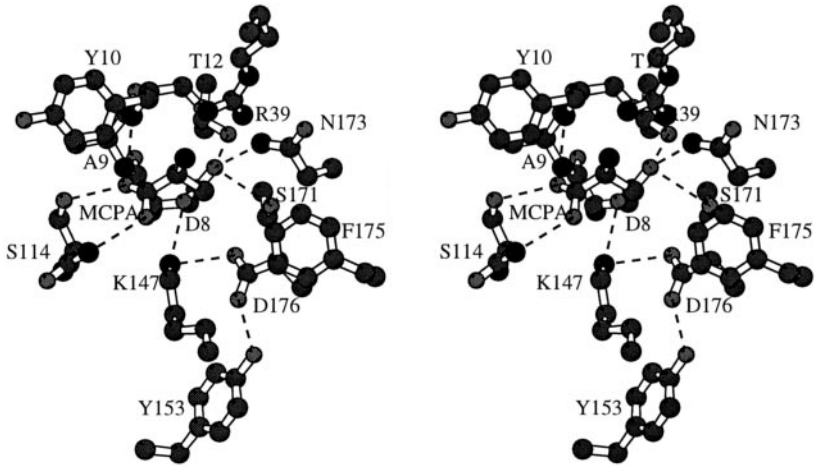

FIG. 7. **Stereo view of the model of L-2-monochloropropionic acid in the active site based on formate position.**

cally important residue contributed by the large atrium subdomain, Arg^{39}, points straight into the active site. Thus, it makes a very likely candidate for halide stabilization and abstraction as it forms a positively charged cradle together with the edges of the phenyl rings of Tyr^{10} and Phe^{175}, which is a Trp in all other L-DEXs. Mutations in both of these aromatic residues significantly decrease the *Pseudomonas* enzyme's activity (9). If we create a structure-based model of a L-2-chloropropionate in the active site with the carboxylate group at the formate position, the chlorine atom fits nicely in this cradle (see Fig. 7).

Halide binding to the guanidinium moiety of arginines and the edges of aromatic side chains has been observed in halorhodopsin (39) and haloalkane dehalogenase (4), respectively. In our structure, two water molecules, W52 and W59, are present at hydrogen-bonding distance from the arginine. The third water molecule (W100) in the active site is found near Asp^8 Oδ1 and water W52. The methyl group of the L-2-chloropropionate model is close to this position, and binding of the substrate would force W100 out of the cavity (see Fig. 7). Note that binding of a D-substrate in a similar way is impossible, as its alkyl group would clash with main chain and side chain atoms of Asp^8, Ala^9, and Tyr^{10}.

In the L-DEX YL enzyme, the orientation of some of the catalytically important residues in the active site is different as well as the solvent structure inside the cavity (see Fig. 8). First of all, the *Xanthobacter* active site is shielded from the solvent, mainly because of the extra small atrium subdomain and the different conformation of the Arg^{39} side chain. It points straight into the cavity, thereby closing the tunnel to the solvent. In the *Pseudomonas* structure, the Arg side chain is rotated by 160° about χ_2, away from the active site, making a long hydrogen bond to Asn^{173} Oδ1. In this way, the halide cradle is disrupted and any interaction with the halogen to be cleaved off is impossible. It might very well be that the L-DEX YL arginine side chain changes its conformation upon substrate binding, orienting itself in a similar way as seen in the *Xanthobacter* structure. Second, in the DhlB structure a formate ion is present. This possibly induces the Ser^{114} side chain to adopt a rotamer state different from the L-DEX YL structure. The orientation of the substrate that we propose on the basis of formate binding deviates from the model of Hisano *et al.* (1), as they suggest binding of the carboxylate group to the corresponding serine and a water molecule. From our structure, it is clear that the serine and three main chain amides are involved. An additional interesting difference is observed for the surroundings of the nucleophile Asp^8. The Oδ2 is bound to Thr^{12} Oγ, Ser^{171} Oγ, and Asn^{173} Nδ2 in DhlB, whereas the *Pseudomonas* dehalogenase only shows a close interaction between the aspartate and the serine. Smaller differences are observed at the bottom of the cleft, where the hydrogen-bonding network of the *Xanthobacter* dehalogenase is made up predominantly by protein-protein interactions, whereas in the *Pseudomonas* enzyme many of the water molecules that are present in its active site cavity are involved.

The two L-2-haloacid dehalogenase structures, the biochemical data available for L-DEX YL, and the model of L-2-chloropropionate binding based on the position of the formate lead us to suggest the following reaction mechanism; a substrate is pulled into the active site by electrostatic forces via the cleft between the main domain and the large atrium subdomain and

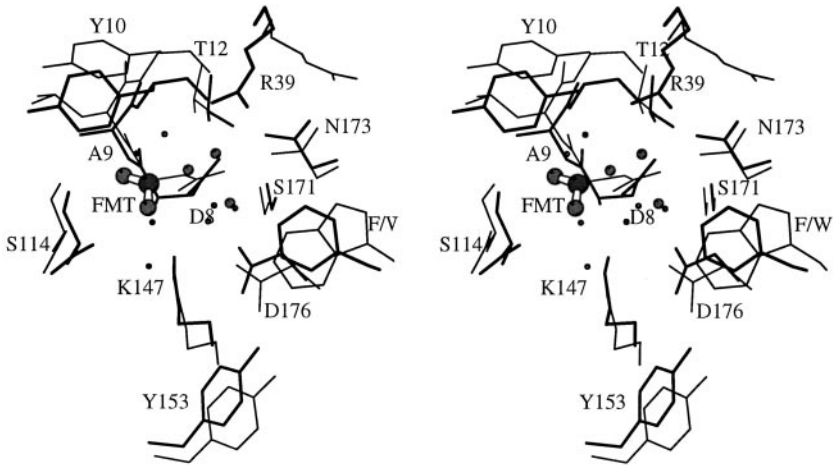

FIG. 8. **Stereo view of the superimposition of dehalogenase active sites of *X. autotrophicus* (*thick lines*) and *Pseudomonas* sp. YL (*thin lines*).** The formate ion and water molecules are shown in ball-and-stick representation.

is bound with its carboxylate group to Ser^{114} and three main chain amide groups. The halogen atom is held in position in a stabilizing halide cradle formed by Arg^{39}, Tyr^{10}, and maybe also Phe^{175}. Subsequently, at the pH optimum of 9–10, the negatively charged Asp^{8} is free to attack the C-2 of the substrate to form an ester intermediate with concurrent cleavage of the carbon-halogen bond. In the next step, the ester is hydrolyzed by a water molecule that attacks on the Cγ of Asp^{8}. The negative charge that develops on the carbonyl oxygen upon formation of the tetrahedral intermediate is stabilized by the oxyanion hole made up by Thr^{12}, Ser^{171}, and Asn^{173}. The L-2-haloacid dehalogenase structures allow us to speculate on the nature and activation of the hydrolytic water molecule. Lys^{147} is the most likely candidate for activation of a water molecule, since its Nζ atom is located in a similar position as the Nϵ2 of the histidine side chain in the catalytic triad of haloalkane dehalogenase and its pK_a is around the relatively high pH optimum of L-2-haloacid dehalogenases. In the DhlB structure with the modeled substrate, however, no water molecules are found within 4 Å from Asp^{8} or Lys^{147}, because all three bound water molecules need to vacate the active site to allow binding of the substrate. It is possible that an as yet unidentified water molecule or hydroxyl ion is imported into the active site after substrate binding or formation of the covalent intermediate, maybe via the smaller tunnel to the active site.

With the detailed structure of the *X. autotrophicus* L-2-haloacid dehalogenase complexed with the substrate-analogue formate, substantial new information about this family of dehalogenases and their substrate binding has become available. It is evident that further research is needed to elucidate the complete reaction pathway of dehalogenation in the L-2-haloacid dehalogenases. In particular, three-dimensional structures of complexes between enzyme and substrates as well as a structure at the optimum pH could help in obtaining more information about the hydrolysis step and the enzyme's stereochemical preferences.

Acknowledgments—We thank Dr. K. S. Wilson, Dr. Gwyndaf Evans, and other staff of the EMBL Outstation (DESY, Hamburg, Germany) for access to the synchrotron data collection facilities and assistance. Work at EMBL Hamburg was supported by the European Union through the HCMP to Large Installations Project under Contract CHGE-CT93-0040.

REFERENCES

1. Hisano, T., Hata, Y., Fujii, T., Liu, J.-Q., Kurihara, T., Esaki, N., and Soda, K. (1996) *J. Biol. Chem.* **271,** 20322–20330
2. Janssen, D. B., Scheper, A., Dijkhuizen, L., and Witholt, B. (1985) *Appl. Environ. Microbiol.* **49,** 673–677
3. Franken, S. M., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1991) *EMBO J.* **10,** 1297–1302
4. Verschueren, K. H. G., Seljée, F., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1993) *Nature* **363,** 693–698
5. Fetzner, S., and Lingens, F. (1994) *Microbiol. Rev.* **58,** 641–685
6. Barth, P. T., Bolton, L., and Thomson, J. C. (1992) *J. Bacteriol.* **174,** 2612–2619
7. Kawasaki, H., Toyama, T., Maeda, T., Nishino, H., and Tonomura, K. (1994) *Biosci. Biotechnol. Biochem.* **58,** 160–163
8. van der Ploeg, J., van Hall, G., and Janssen, D. B. (1991) *J. Bacteriol.* **173,** 7925–7933
9. Kurihara, T., Liu, J. Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., and Soda, K. (1995) *J. Biochem.* (*Tokyo*) **117,** 1317–1322
10. Benning, M. M., Taylor, K. L., Liu, R.-Q., Yang, G., Xiang, H., Wesenberg, G., Dunaway-Mariano, D., and Holden, H. M. (1996) *Biochemistry* **35,** 8103–8109
11. McConnell, G., Ferguson, D. M., and Pearson, C. R. (1975) *Endeavour* **34,** 13–18
12. Liu, J.-Q., Kurihara, T., Miyagi, M., Esaki, N., and Soda, K. (1995) *J. Biol. Chem.* **270,** 18309–18312
13. Ridder, I. S., Rozeboom, H. J., Kingma, J., Janssen, D. B., and Dijkstra, B. W. (1995) *Protein Sci.* **4,** 2619–2620
14. Messerschmidt, A., and Pflugrath, J. W. (1987) *J. Appl. Crystallogr.* **20,** 306–315
15. Kabsch, W. (1988) *J. Appl. Crystallogr.* **21,** 916–924
16. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S. S., eds) pp. 56–62, SERC Daresbury Laboratory, Warrington, United Kingdom
17. Furey, W., and Swaminathan, S. (1997) *Methods Enzymol.* **277,** 590–620
18. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119
19. Navaza, J. (1994) *Acta Crystallogr. A* **43,** 157–163
20. Brünger, A. T. (1992) *Nature* **355,** 472–475
21. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science* **235,** 458–460
22. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D* **53,** 240–255
23. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D* **49,** 129–147
24. Read, R. J. (1986) *Acta Crystallogr. A* **42,** 140–149
25. Bhat, T. N. (1988) *J. Appl. Crystallogr.* **21,** 279–281
26. Vellieux, F. M. D., and Dijkstra, B. W. (1997) *J. Appl. Crystallogr.* **30,** 396–399
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
28. Nicholls, A., and Honig, B. (1993) GRASP, Columbia University, New York
29. Eisenhaber, F., and Argos, P. (1993) *J. Comput. Chem.* **14,** 1272–1280
30. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. D* **50,** 178–185
31. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D* **50,** 760–763
32. Luzzati, V. (1952) *Acta Crystallogr.* **6,** 142–152
33. Engh, R. A., and Huber, R. (1991) *Acta Crystallogr A* **47,** 392–400
34. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature* **250,** 194–199
35. Liu, J. Q., Kurihara, T., Hasan, A. K. M. Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., and Soda, K. (1994) *Appl. Environ. Microbiol.* **60,** 2389–2393
36. Verschueren, K. H. G., Franken, S. M., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1993) *J. Mol. Biol.* **232,** 856–872
37. Taylor, K. L., Xiang, H., Liu, R. Q., Yang, G., and Dunaway-Mariano, D. (1997) *Biochemistry* **36,** 1349–1361
38. Liu, J.-Q., Kurihara, T., Miyagi, M., Tsunasawa, S., Nishihara, M., Esaki, N., and Soda, K. (1997) *J. Biol. Chem.* **272,** 3363–3368
39. Rüdiger, M., Haupts, U., Gerwert, K., and Oesterhelt, D. (1995) *EMBO J.* **14,** 1599–1606
40. Esnouf, R. M. (1997) *J. Mol. Graphics* **15,** 133–138
41. Kraulis, P. (1991) *J. Appl. Crystallogr.* **24,** 946–950